

OCT-NET: A CONVOLUTIONAL NETWORK FOR AUTOMATIC CLASSIFICATION OF NORMAL AND DIABETIC MACULAR EDEMA USING SD-OCT VOLUMES

Oscar Perdomo¹, Sebastian Otálora², Fabio A. González¹, Fabrice Meriaudeau³, Henning Müller²

¹ MindLab Research Group, Universidad Nacional de Colombia, Colombia,

² University of Applied Sciences Western Switzerland (HES-SO), Switzerland,

³ Universiti Teknologi PETRONAS, Malaysia.

ABSTRACT

Diabetic macular edema (DME) is one of the most common eye complication caused by diabetes mellitus, resulting in partial or total loss of vision. Optical Coherence Tomography (OCT) volumes have been widely used to diagnose different eye diseases, thanks to their sensitivity to represent small amounts of fluid, thickness between layers and swelling. However, the lack of tools for automatic image analysis for supporting disease diagnosis is still a problem. Convolutional neural networks (CNNs) have shown outstanding performance when applied to several medical images analysis tasks. This paper presents a model, OCT-NET, based on a CNN for the automatic classification of OCT volumes. The model was evaluated on a dataset of OCT volumes for DME diagnosis using a leave-one-out cross-validation strategy obtaining an accuracy, sensitivity, and specificity of 93.75%.

Index Terms— Diabetic macular edema, OCT, convolutional neural network, deep learning, eye disease diagnosis.

1. INTRODUCTION

Diabetic Macular Edema (DME) is one of the main complications related to Diabetes, and the largest leading cause of partial and total loss of vision [1]. DME affects central vision, leading to distortion of part of the visual field and blurred vision, even limiting the ability of people to do activities, such as reading, driving or walking [2]. One of the main problems regarding DME diagnosis is its occurrence at any stage of Diabetic Retinopathy (DR), so a precise diagnosis is critical for the correct treatment, avoiding health complications and associated costs [1, 2].

DME diagnosis using eye fundus images is performed by experts looking for signs of retinopathy, retinal thickening and presence of exudates in the macula and the fovea [3, 4]. One of the main limitations of using eye fundus images for diagnosis is the subjectivity of the analysis and strong disagreement between experts on the grade of the disease [5].

Spectral domain OCT (SD-OCT) is a common noninvasive imaging technique useful to make DME diagnosis more objective and reliable [4]. The diagnosis is performed

measuring five patterns of structural changes as follows: diffuse retinal thickening (DRT), cystoid macular edema (CMD), serious retinal detachment (SRD), posterior hyaloidal traction (PHT) with/without tractional retinal detachment (TRD) [3], [5].

Machine learning techniques have been applied to DME classification in OCT scans. Alsaih et al. combine the extraction of histograms of oriented gradients with local binary patterns (LBPs) [6] in order to detect DME volumes. Lu et al. [7] propose a method for AMD diagnosis detecting the B-scan frames containing fluid regions and converting each slice into a bag of features. Methods based on deep learning also reported good performance in classification and segmentation tasks. Roy et al. [8], propose a fully convolutional deep architecture (ReLayNet) with a joint loss function for the segmentation of retinal layers in OCT scans. Lee et al. developed an automated segmentation based on convolutional neural networks (CNN) that detect DME [9] and Age-related macular degeneration (AMD) [10] on OCT volumes. Furthermore, Karri et al. [11] presented a deep learning approach that fine-tunes a pre-trained GoogLeNet on a dataset of OCT images with DME, with dry AMD and normal cases.

A deep learning approach that combines a pre-trained VGG with ensemble classifiers for DME volume classification was studied [12]. Although first deep learning work was focused on the fine-tuning of pre-trained weights of known CNN architectures, the design of new end-to-end CNNs to perform an accurate diagnosis of DME OCT volumes has not been explored, yet.

This paper presents an end-to-end CNN for automatic DME diagnosis using SD-OCT volumes with an accuracy of $93.75 \pm 3.125\%$, and, a sensitivity, and specificity of 93.75%, improving results in accuracy and specificity compared to a baseline method. The remainder of this paper is organized as follows: an overview of our proposed method is presented in Section 2. The experimental setup is described in Section 3. The experimental results are reported in Section 4. Finally, Section 5 draws the conclusions of our experiments.

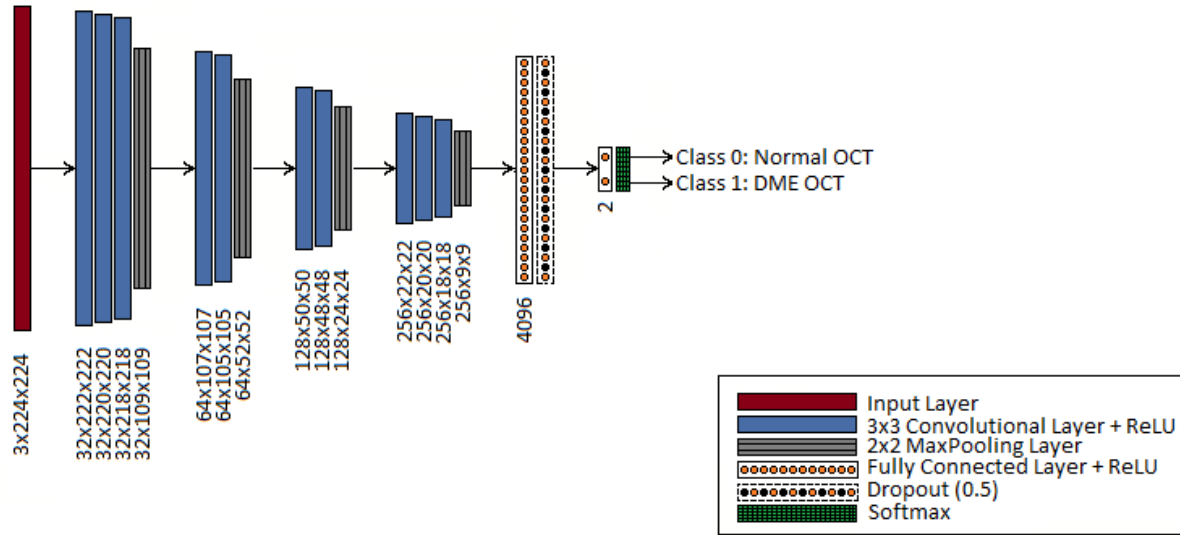


Fig. 1. Architecture of the OCT-NET model. The input layer receives every scan with a resolution of 224×224 and is stacked twice to create a 3-channel image. The number of feature maps and the size are reported below every layer.

2. METHODS

The OCT-NET is an end-to-end deep learning model composed of 2 main blocks: the first contains 4 sub-blocks with convolutional and max-pooling layers with a different number of filters to extract features learned, and the last block has two fully connected and one dropout layers to perform the scan classification among OCT volumes. An overall view of the model and its parameters are shown as block diagram in Fig. 1 and described in detail in Section 2.2.

2.1. Cropping and resizing OCT volumes

Cropping and resizing images are common pre-processing operations applied to the input of convolutional neural network models for medical image analysis. The goal is to remove irrelevant information of the image for extracting the Region of Interest (RoI) and the aspect ratio of the cropped image keeping the same as the original image by applying a set of transformations to the volumes. Noise removal algorithms implemented in the baseline method were not taken into account in our approach.

The RoI of an OCT can be cropped by a manual segmentation by medical experts or automatically detected by a computer-aided detection system. However, the location of layers inside an OCT scan oscillates among the top, middle or bottom position, making it difficult to perform an automatic detection. Because of this, a median filter on every scan was implemented in order to highlight and ease the extraction of relevant information through layers [13]. Hence, RoIs can be extracted easily by taking the bounding box of the nine layer boundaries on retinal optical coherence tomography by scan.

Specifically, bounding boxes ranged between 512×512 and 300×512 pixels. Finally, bounding box images were centered without scaling and preserving the surrounding region with a resolution of 224×224 and every scan was stacked twice to handle an input size of $224 \times 224 \times 3$.

2.2. OCT-NET deep learning model

The OCT-NET is a 12 layer CNN model that receives as an input scans with a resolution of $224 \times 224 \times 3$ to classify an OCT volume. Our method is based on 4 sub-blocks that contain ten convolutional layers with an incremental number of filters from 32 to 256 in order to extract and learn different data-representations. The model is trained using stochastic gradient descent as usual.

A convolutional layer contains a set of learned filters that convolve with the input generating an activation map for each filter. Convolutional layer outputs are stacked as inputs for a max-pooling layer applying a non-linear size reducer to the activation choosing the maximum value of a set of contiguous pixels. A dropout layer consists of random deactivation of a fraction of input units at each update during training time (0.5 in our method) [14]. Finally, two fully-connected layers connect all the activated neurons in the previous layer to the next layer with 4096 and 2 units respectively.

The parameters used are as follows: ten convolutional layers with kernel size of 3×3 and stride of 1×1 , three max-pooling layers with pool size of 2×2 and stride of 2×2 , one dropout layer, and two fully-connected layers.

3. EXPERIMENTAL SETUP

3.1. SERI Dataset

The Singapore Eye Research Institute (SERI) database contains 32 Spectral Domain-Optical Coherence Tomography (SD-OCT) volumes acquired by SERI following the requirements according to the ethic committee [12]. The database is composed of 16 normal and 16 DME OCT volumes. Each volume contains 128 B-scans or cross-sectional scans with a resolution of $1,024 \times 512$ px as shown in Fig. 2.

OCT volumes were captured with a CIRRUS TM SD-OCT manufactured by Carl Zeiss Meditec. OCT volumes were analyzed and labeled by experts as normal or DME volumes, according to findings among the layers, retinal thickening, location of exudates and hard-exudates, intraretinal cystoid space information, and subretinal fluid. The dataset was cropped and resized (keeping the aspect ratio) to a dimension of $224 \text{ px} \times 224 \text{ px}$ as discussed in Section 2.1.

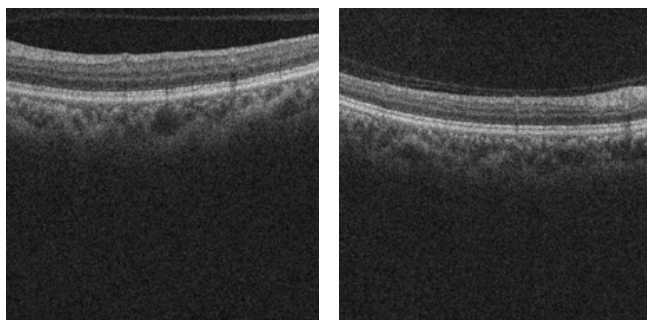


Fig. 2. [Left] Normal OCT scan and [Right] DME OCT scan with diffuse thickening of the outer and middle layers with losses of the regularity of their layers.

3.2. Evaluation

The proposed CNN model was trained using the RMSprop optimizer with a decay term of commonly used value $\rho = 0.9$. The batch size and learning rate were explored with a grid search and are reported in Table 1. The results were stable for various numbers of epochs. It was experimentally set to 5 as it obtained high accuracy while limiting the training time that is on average to 5 minutes per fold.

A deep learning approach combined with ensemble classifiers was chosen as baseline [12]. In this work, a pre-trained VGG-16 CNN was used and every OCT volume of 128 scans that was triplicated to handle an input size of $224 \times 224 \times 3$. The feature vectors from three fully-connected layers: the first two (4096 units) and the last one (1000 units), were classified using two KNNs (with $K = 1$ and 3), and one Random Forest (100 trees) classifier. The best performance obtained was 93% in accuracy, 87% in sensitivity and 100% in specificity.

The proposed method uses two considerations reported for the baseline method [12] to classify an OCT volume as follows: the SERI dataset was split into 32 independent folds using k-fold cross validation (leave-one-patient-out), where one fold contains 31 volumes in the training set and one volume to test the model. The classification of one OCT volume was performed using a quorum rule, where if 65 or more scans were classified as class 0, the volume was labeled to class 0.

The OCT-NET approach was implemented with Keras using a Theano backend on a GeForce GTX TITAN X from NVIDIA. The loss and accuracy were monitored on the training data and the best model was saved and assessed on the patient left out.

4. RESULTS

We analyzed the performance of the proposed model with a systematic exploration of parameters applied to all OCT volumes using k-fold cross validation (leave-one-patient-out) according to Section 3.2. A systematic exploration of learning rate (lr) and batch size (bs) is reported in Table 1. The best performance of the proposed model was obtained with a learning rate of 0.00001 and a batch size of 16.

Table 1. Performance measures of the proposed method on the test data, bold values show the best score among all methods.

Method	bs	lr	Acc.	Sens.	Spec.
OCT-NET	8	0.001	53.12%	51.85%	60%
		0.0001	56.25%	53.57%	75%
		0.00001	87.5%	83.33%	92.86%
	16	0.001	46.87%	48.39%	0%
		0.0001	56.25%	53.33%	56.25%
		0.00001	93.75%	93.75%	93.75%

Finally, we compared the performance of the OCT-NET with and without dropout layer against the baseline method. An area under the ROC curve of 0.927 was obtained using the proposed model. The accuracy, sensitivity and specificity are presented in Table 2. The proposed method clearly outperforms Awais et al. [12], in accuracy and sensitivity.

5. DISCUSSION AND CONCLUSION

We present a new end-to-end OCT classification model based on CNNs that is able to effectively detect the relevant patterns for diabetic macular edema using raw images. The experimental results show that the proposed method is highly effective for classifying OCT volumes as normal or DME (see Table 2). The results improve the accuracy and sensitivity over the baseline due to the ability of the end-to-end method to learn relevant information along layers of its structure.

Table 2. Performance measures for the baseline model with raw images without and with cropping (dataset 1 and 3), the proposed method and the proposed method without dropout, bold values show the best score among all methods.

Method	Acc.(%)	Sens.(%)	Spec.(%)
Dataset 1 [12]	93	87	100
Dataset 3 [12]	87.5	93.5	81
Proposed method	93.75±3.125	93.75	93.75
without dropout	81.25	77.78	85.71

Additionally, the dropout layer in our approach acts like a multiple-ensemble approach testing several configurations during training and being more efficient than combining transfer learning and ensemble learning methods, as reported by Awais et al. [12].

The proposed method is an option to use for other ocular diseases, such as age-related macular degeneration or diabetic retinopathy, optimizing work-time without the need for segmentation and measuring of layers and abnormalities as traditionally processed by experts classification.

6. ACKNOWLEDGEMENT

Oscar Perdomo thanks COLCIENCIAS and HES-SO for funding this research with a doctoral grant and an international internship respectively. We appreciate the efforts devoted by Carol Cheung and Tien Y Wong from the Chinese University of Hong Kong (CUHK) and Singapore Eye Research Institute (SERI) to collect SD-OCT volumes.

7. REFERENCES

- [1] Gabriele E Lang, “Diabetic macular edema,” *Ophthalmologica*, vol. 227, no. Suppl. 1, pp. 21–29, 2012.
- [2] Oscar Perdomo, Sebastian Otalora, Francisco Rodríguez, John Arevalo, and Fabio A González, “A novel machine learning model based on exudate localization to detect diabetic macular edema,” *OMIA 2016, Held in Conjunction with MICCAI 2016*, pp. 137–144, 2016.
- [3] Lihteh Wu, Priscilla Fernandez-Loaiza, Johanna Sauma, Erick Hernandez-Bogantes, and Marissé Masis, “Clasificación de diabetic retinopathy and diabetic macular edema,” *World journal of diabetes*, vol. 4, no. 6, pp. 290, 2013.
- [4] Jie Ding and Tien Yin Wong, “Current epidemiology of diabetic retinopathy and diabetic macular edema,” *Current diabetes reports*, vol. 12, no. 4, pp. 346–354, 2012.
- [5] Rui Bernardes and José Cunha-Vaz, *Optical coherence tomography: a clinical and technical update*, Springer Science & Business Media, 2012.
- [6] Khaled Alsaih, Guillaume Lemaitre, Mojdeh Rastgoo, Joan Massich, Désiré Sidibé, and Fabrice Meriaudeau, “Machine learning techniques for diabetic macular edema (dme) classification on sd-oct images,” *Biomedical engineering online*, vol. 16, no. 1, pp. 68, 2017.
- [7] Donghuan Lu, Weiguang Ding, Andrew Merkur, Marinko V Sarunic, and Mirza Faisal Beg, “Multiple instance learning for age-related macular degeneration diagnosis in optical coherence tomography images,” in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 139–142.
- [8] Abhijit Guha Roy, Sailesh Conjeti, Sri Phani Krishna Karri, Debodoot Sheet, Amin Katouzian, Christian Wachinger, and Nassir Navab, “Relaynet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional network,” *arXiv preprint arXiv:1704.02161*, 2017.
- [9] Cecilia S Lee, Ariel J Tying, Nicolaas P Deruyter, Yue Wu, Ariel Rokem, and Aaron Y Lee, “Deep-learning based, automated segmentation of macular edema in optical coherence tomography,” *bioRxiv*, p. 135640, 2017.
- [10] Cecilia S Lee, Doug M Baughman, and Aaron Y Lee, “Deep learning is effective for classifying normal versus age-related macular degeneration oct images,” *Ophthalmology Retina*, vol. 1, no. 4, pp. 322–327, 2017.
- [11] Sri Phani Krishna Karri, Debjani Chakraborty, and Jyotirmoy Chatterjee, “Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration,” *Biomedical optics express*, vol. 8, no. 2, pp. 579–592, 2017.
- [12] Muhammad Awais, Henning Müller, and Fabrice Meriaudeau, “Classification of sd-oct images using deep learning approach,” in *Signal and Image Processing Applications (ICSIPA 2017), 2017 IEEE International Conference on*. IEEE, 2017.
- [13] Ananda R Natarajan, G Mohandass, and G Hari Krishnan, “Comparative analysis of optical coherence tomography retinal images using multidimensional and cluster methods,” *Biomedical Research*, vol. 26, no. 2, 2015.
- [14] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.